

Experiment Design & Results Template

Purpose of this template?

A design brief for an A/B experiment defines the test's objectives, hypotheses, and key design elements. It ensures clarity on the variables being tested, the target audience, and success metrics. This document aligns stakeholders, streamlines execution, and serves as a reference for analysing results to make data-driven decisions.

[Create A Hypothesis](#)

[Perform A Power Analysis: Baseline Data](#)

[Provide Experiment Information](#)

[Design References](#)

[Capture Experiment Results](#)

[Experiment Results & Learnings](#)

Create A Hypothesis

What hypothesis are you trying to validate with this experiment?

Need help with writing a hypothesis? Visit our Help Center:

<https://help.split.io/hc/en-us/articles/360055681831-Constructing-a-Hypothesis>

Problem - What?	Customers don't know that hotel rooms include many additional amenities and benefits
Strategy - How?	Create awareness
Tactics - If?	Automatically show detailed content (that is normally hidden) under the room type description
Result - Then?	Increased bookings
Hypothesis: <i>By automatically showing detailed content under the room type description to create awareness, this should result in increased bookings, because currently customers don't know that hotel rooms include many additional amenities and benefits.</i>	

Perform A Power Analysis: Baseline Data

[Note: only necessary when using our Fixed Horizon testing method]

<https://speero.com/ab-test-calculator>

<p>Weekly Traffic</p> <p><i>Volume of traffic to the test page per week</i></p>	
<p>Weekly Conversions</p> <p><i>Weekly conversions from your primary success metric. Use uniques or totals depending on what your success measure is.</i></p>	
<p>Number of Treatments</p> <p><i>The number of treatments you'd like to experiment with. This will be a variable that will directly impact run time for statistical significance depending on the traffic and current conversion rate on the page with the test change</i></p>	

Visit our [Help Center](#) for more details regarding performing a power analysis prior to running experiments:

<https://help.split.io/hc/en-us/articles/360034040851-Sample-size-and-sensitivity-calculators>

<p>Sample size</p> <p><i>How many <traffic type> do you expect to be included in the test given the targeting criteria and the intended run time?</i></p>	4,570
<p>Minimum detectable effect of key metric</p> <p><i>What is the result of the power analysis for the key metric? How large does the impact need to be in order to reach statistical significance?</i></p>	10%

Experiment Settings <i>What will you use for the experiment settings? What is the minimum for calculating statistical significance? What will you use for the significance threshold (p-value)? Will you apply multiple comparison corrections (recommended)?</i>	Minimum sample size: 10 Significance threshold (p-value): 0.05 Experimental review period: 7 days Multiple comparison correction: Yes
Expected Run Dates & Duration <i>How long will you run the test? (e.g. 7 days, 2 weeks, etc.) When do you expect to run the experiment?</i>	Run time: 2 weeks Launch date: 1/1/99

Provide Experiment Information


Experiment name <i>What is your experiment called?</i>	front_end_show_more_info
Type of Experiment <i>Optimisation, A/B/n, Validation/Do No Harm, Painted Door Research, Backend changes, Progressive rollout, UX</i>	Optimisation (A/B)
Experiment URL <i>What is the URL of the experiment configuration?</i>	[add platform URL]
Status <i>What is the status of this project?</i> PRE-PRODUCTION/ QA & INTERNAL TESTING / RAMPING / EXPERIMENTING /100% RELEASED PERMANENT /KILLED / REMOVED FROM CODE	Experimenting
Experiment Owner <i>Who is a good point of contact for more information?</i>	Heather Davis
Description <i>What is being changed, or what new feature are you testing?</i>	On the room details page, show/hide room and amenities information
Location of Experiment/Point of Exposure <i>Does it exist on a certain page / area of the product?</i>	Split Hotel Training Reservations Page
Treatment(s) <i>What are the variations of software treatments?</i>	<ul style="list-style-type: none"> ● Treatment A: show_text <ul style="list-style-type: none"> ○ New: Show detailed room info by default ● Treatment B: dont_show_text <ul style="list-style-type: none"> ○ Current: have to click "more info" to see detailed room info <p><i>See designs below</i></p>

<p>Primary Key Metric(s)</p> <p><i>Which metric(s) will be impacted?</i></p> <p><i>What is your Overall Evaluation Criteria?</i></p>	<p>Bookings per User</p>
<p>Secondary Metric(s)</p> <p><i>Which metric(s) are most likely to impact?</i></p>	<ul style="list-style-type: none"> ● Booking Avg Price per User ● Bookings per Platinum User ● Revenue per User ● Completed Bookings per User ● Sum of Bookings per User ● Bounce Rate per User ● Scroll Depth per User ● Hotel Content Clicks per User ●
<p>Guardrail metric(s)</p> <p><i>What are you monitoring to ensure it has not been impacted?</i></p>	<ul style="list-style-type: none"> ● Average Page Load ● Average Session Length per User
<p>Traffic Type and Allocation</p> <p><i>What traffic type is being used for the experiment?</i></p> <p><i>e.g., account, if all users in an organisation should see the same treatment</i></p>	<ul style="list-style-type: none"> ● Traffic Type: Users ● Allocation: 100%
<p>Targeting Criteria: Rules, Segments</p> <p><i>Who will be included in the test? Are there considerations of cohorts or audiences? (e.g. account types, user designations)</i></p>	<ul style="list-style-type: none"> ● Targeting Rule(s): Default (50/50 split) ● Segment(s): None ● Individual Target(s): None
<p>Approvals</p> <p><i>Who should be approving and launching your experiment?</i></p>	<p>None</p>
<p>Jira Ticket</p> <p><i>Link to Jira url</i></p>	<p>[Add Jira ticket]</p>
<p>Other notes and considerations</p> <p><i>List any particular points of worry, especially in terms of circumstances that would entail shutting down the experiment, or considerations of other stakeholders or business owners.</i></p>	<p>Keep in mind any holiday/weekend/seasonal traffic, manage stakeholder expectations for possible outcomes and learnings</p>

Design References

<p style="text-align: center;">Treatment A:</p> <p style="text-align: center;"><i>Add a screenshot for reference</i></p>	<p style="text-align: center;">Treatment B:</p> <p style="text-align: center;"><i>Add a screenshot for reference</i></p>

Capture Experiment Results

Metrics Scorecard				
Metric Name <i>Overall Evaluation Criteria</i>	<Metric Value A> <i>Baseline or Control</i>	<Metric Value B> <i>Variation</i>	$\Delta\%$ <i>Increase/ Decrease</i>	Significance  <i>p-value</i> <i>Desired, Undesired, or Inconclusive?</i>
Key Metric: <Name>				
Key Metric: <Name>				
Supporting Metric: <Name>				
Supporting Metric: <Name>				
Guardrail Metric: <Name>				
Guardrail Metric: <Name>				

Metric Export PDF	
<i>Attach a copy of the results exported from the Split app through "Share results" button</i>	
Export Date	mm/dd/yy hh:mm
Targeting Rule	Default
Comparison treatment	Treatment A: show_text
Compared against baseline treatment	Treatment B: dont_show_text

Show_More_Info

[Edit experiment](#) [Settings](#) ⋮

[Health Check](#) [Assignment source](#) [Hypothesis](#) Owners: Administrators

Scope Starts at: 02/07/2025 12:13AM Ends at: 02/21/2025 12:13AM Rule: default rule 119,720 exposures



Comparison treatments show_text X × ▾

[Recalculate all metrics](#)

Key metrics [Manage key metrics](#) Last calculated 02/21/2025 12:13AM [↻](#)

Metric	Treatment	Direction	Impact	P-value
Total Booking Dollars Per User	■ dont_show_text	-	-	
	■ show_text	Desired	13.15% (±8.85%)	<0.001
Booking Avg Price per User	■ dont_show_text	-	-	
	■ show_text	Inconclusive ⓘ	-0.23% (±0.85%)	1.000
Bookings per User	■ dont_show_text	-	-	
	■ show_text	Desired	2.47% (±1.98%)	<0.001

Guardrail metrics Last calculated 02/21/2025 12:13AM [↻](#)

Metric	Treatment	Direction	Impact	P-value

Experiment Results & Learnings

Highlights and Learnings

What did you learn? What are the key take aways? Other notes to share? Was there any significant degradation seen in any of the metrics (primary, supporting or guardrail)?

- Notes
- Notes

Next Steps

Did the results pass the decision criteria? Will this change/feature be rolled out to all users? What knowledge was gained that suggests future experiments in this area?

- Notes
- Notes